

Dimensions of parallelism: some policy applications of experimental methods

CHARLES R. PLOTT

7.1 Introduction

The term “parallelism” refers to a vague notion about how observations of simple laboratory phenomena can help one understand and predict the behavior of a complicated and changing world. Of what use are experimental results to someone who is interested in something vastly larger and more complicated, perhaps fundamentally different than anything that can be studied in a laboratory setting? Questions such as this and the related notion of parallelism have probably existed from the earliest development of scientific experimental methodology, and although I found the term in a paper by Vernon Smith (1980) the notion itself pervades all branches of science and engineering.

The purpose of this chapter is to isolate some examples of how the issue of parallelism has been approached in economics. The chapter outlines several strategies that have been employed in attempts to use experimental research in actual policy decision making. The topic to be explored is how issues have been posed in these policy-related studies so that experimental methods could be applied. The discussion is limited to 10 instances in which I have been involved personally at some level.

Many different opinions exist about experimental methodology and the relationship between laboratory work, field studies, and policy decisions. The opinions are strongly held and are just as likely to be held by those with no experience at all in applying the methods as by those with much experience. For example, the textbook by Samuelson and Nordhaus (1983, p. 8) boldly claims that experiments in economics

The financial support of the National Science Foundation and the California Institute of Technology Program for Enterprise and Public Policy is gratefully acknowledged. I also wish to thank Alvin Roth and Howard Kunreuther, who provided comments on an early draft.

are impossible. Presumably these authors believe that some sort of field study is the only way to approach an application of experimental methods. Referee reports frequently reflect methodological philosophies and related concepts of parallelism. Every experimentalist who has submitted a paper to a professional journal has read a referee report aggressively claiming that the experiments had nothing to do with the "real world" or that the experiments were not "relevant" for some reason or another. My impression is that such critics have very narrow views about the connections between laboratory and naturally occurring situations, and they approach experimental methods with unrealistic expectations about what can be learned from applications.

This impression brings me to my point: Economists should keep an open mind about experimental methodology and should judge work by the statements of results rather than by methodological principles. Methodological principles should evolve from our experiences with what works and what does not work. That point is reflected in the title and organization of this chapter. The topic is policy research as opposed to basic research. The issues are: What was attempted, what seemed to work and why, what was a flop and why?

The examples are organized according to what seems to have been the principal strategy for using the experiments. Each strategy can be viewed as a "dimension" or form of parallelism between policy problems and laboratory experiments. Five different strategies are identifiable. Each section treats a different strategy. The discussion includes a general description of the strategy, the context of the policy problem, and the role of the experiments in the final policy decision if any decision resulted.

7.2 Ex post evaluation of a decision: the flying club

A policy decision was made. An action was taken. A result was observed. What influence did the policy have on the observed result? The question suggests a possible role for experiments in the ex post evaluation of a policy decision.

The ex post evaluation of a policy decision motivated the agenda experiments reported in Levine and Plott (1977). The policy decision in this case was one that Mike Levine and I made to promote the use of a particular agenda by a large flying club that was selecting a fleet of aircraft. Unknown to all members of the club but ourselves, the agenda was designed to influence the club to choose the fleet of aircraft most preferred by Levine. The theory underlying the design of the agenda was untested at the time the decision was made, and the preferences of

the members of the club were largely unknown. Nevertheless, the club used the agenda and chose the fleet the agenda was designed to secure. Was the agenda responsible for the group choice? Was the "policy" a success? The question can never be answered, but educated guesses can be made. The role of the experiments is to provide the foundation for making an educated guess.

The context of the decision will make the methodological issue clear. The flying club had a fixed amount of money to spend on aircraft. It had many options, including a variety of makes, sizes, and equipment. The club could buy several aircraft, each of which could be different, and the number of possible fleets was in the thousands. The preferences of the members of the club differed substantially. Since there were many possibilities and no unanimous opinion, how was the group to decide?

The agenda used by the club is shown in Figure 7.1. The first question addressed the primary or basic fleet to be purchased. After a primary fleet was chosen, the next issue was the number of aircraft. After that question was resolved, the group addressed the question of whether more than one type of aircraft should be purchased and how many. The final question was how the aircraft should be equipped.

Notice that each question partitions the options into two sets: a set to be rejected and a set to keep for further consideration. In order to understand this point consider Figure 7.2. The letter E represents a Bonanza E; F a Bonanza F; C a Cessna; and A a Bonanza A. For practical purposes after the basic fleet was decided, the remaining options were considered by a sequence of questions. First, the number of aircraft to be purchased was considered (six-plane fleet vs. seven-plane fleet). After that a series of questions addressed the composition of the fleet. Notice that the agenda can be represented in a tree form as shown and that a reordering or rewording of the agenda would lead to the formation of a different tree.

The agenda represented in Figure 7.2 was chosen to induce the group to choose the option EEEFFCC, which we preferred. This was a fleet of five small Bonanza Es and Bonanza Fs with two larger Cessnas. The basic idea was to use conflicts among members of the group and the majority rule to eliminate options at each stage of the agenda until the remaining choices would result in the option (EEEFFCC) that we wanted. Each different stage of the agenda uses a different majority to eliminate options that we wanted eliminated.

Our reasoning in designing this agenda involved aspects of theory that had not been tested directly, so the utility of our project could certainly be questioned. The agenda seemed to work according to plan, but how do we know? Perhaps the group's choice was fortuitous and had nothing at all to do with the agenda.

This is the agenda for the Group III equipment meeting. Your subcommittee has tried to define a series of problems facing the group and to give you an opportunity to express your preferences in resolving them. We suggest that you take a few moments to look over this agenda and familiarize yourselves with the choices facing us, then come to the meeting, participate in the discussion, and vote by show of hands on the choices presented in alternatives 2-6. We would like to present the Board with the most comprehensive possible expression of Group III opinion. Please come.

1. INTRODUCTION:

Availability, Type variety, Previous Depreciation problems, Needs of the Group vs. Cost, Safety, Radio Equipment.

2. PRIMARY AIRCRAFT TYPE:

PROBLEM: Survey suggests that many Group III members prefer that the main part of the group fleet be four-seat Bonanzas. Should these be all the same age? If so, we could sell all existing Bonanzas and buy new F-33A's or we could sell only the V and F and buy used E-33A's. If they can be different ages, should we keep our E's and add new F's? Or do we want C-210's? Previous depreciation practices may affect these choices.

INPUTS:

Costs and rates for new F-33A's and refurbished E-33A's.

Depreciation problems.

Maintenance comparisons.

Availability and price of used aircraft.

VOTE: PRIMARY FLEET TYPE SHOULD BE:

- a. All new F-33A's at about \$29.00 hour;
- b. Refurbished E-33A's at about \$24.00 hour;
- c. Mixed new F-33A's at about \$28.00 hour and refurbished F-33A's at about \$24.00 hour;
- d. New C-210's at about \$25.00 hour.

3. SIZE OF GROUP III FLEET:

PROBLEM: Survey suggests that membership considers present availability to be unsatisfactory. This summer we operated with a little over five aircraft available. We have based our rates on 500 hrs./yr./aircraft. With only five aircraft available, we are flying more than that. We can clearly operate six aircraft at 500 hrs./yr./aircraft. We might be able to operate seven at that rate. We almost certainly couldn't operate eight at 500. If we assume, conservatively, that a seventh aircraft would operate 400 hours and an eighth 300 hours, the question becomes, "how much availability do we want to pay for?"

INPUTS:

Alternative ways of paying for availability.

VOTE:

Cost increases associated with availability.

- a. 6
- b. 7
- c. 8

Figure 7.1. Flying club agenda. From Levine and Plott (1977).

SHOULD THE FLEET INCLUDE AIRCRAFT OTHER THAN THE PRIMARY TYPE?

PROBLEM: Most members indicated an occasional need for a five- or six-place airplane. Others indicated a desire to fly aircraft other than Bonanzas. There are advantages in scheduling, rate uniformity, majority choice, and type familiarity in keeping the fleet homogeneous. The advantages of operating more than one type include optimizing for different mission requirements and accommodating minority preferences.

INPUT:

Safety aspects of mixed fleets.

Survey input on desire for 5-place, 6-place, and mixed fleet.

VOTE: FLEET SHOULD BE:

- a. All primary type;
- b. Mixture of mostly primary type and some six-place.

IF SOME SIX-PLACE SHOULD BE INCLUDED, SHOULD THEY BE BONANZA A-36's or C-210's?

PROBLEM: Each of the two has advantages and disadvantages and different costs.

INPUT:

Weight and Balance and Performance comparisons.

Maintenance comparisons.

A-36 costs and advantages.

C-210 costs and advantages.

VOTE: SHOULD SECONDARY AIRPLANES BE:

- a. A-36 at about \$31.50 hour?
- b. C-210 at about \$27.00 hour?

ADDITIONAL EQUIPMENT.

PROBLEM: It has been club policy (and probably will be in the future) to equip aircraft alike. Most of the group has indicated a preference for glideslopes, and the cost discussions so far have included them. Others have discussed DME's, radio-coupled autopilots (no altitude hold), and encoding altimeters (to meet Group I TCA requirements starting 7/1/74).

INPUT:

Cost and uses of equipment. Increase in cost/hour.

VOTE:

Would you like to have the following equipment if it increased cost per hour by the following amounts?

YES NO

- DME at about \$—hour.
- Coupled autopilot at about \$—hour.
- Encoding altimeter at about \$—hour.

SUMMARY AND RECOMMENDATIONS TO BE MADE TO THE BOARD

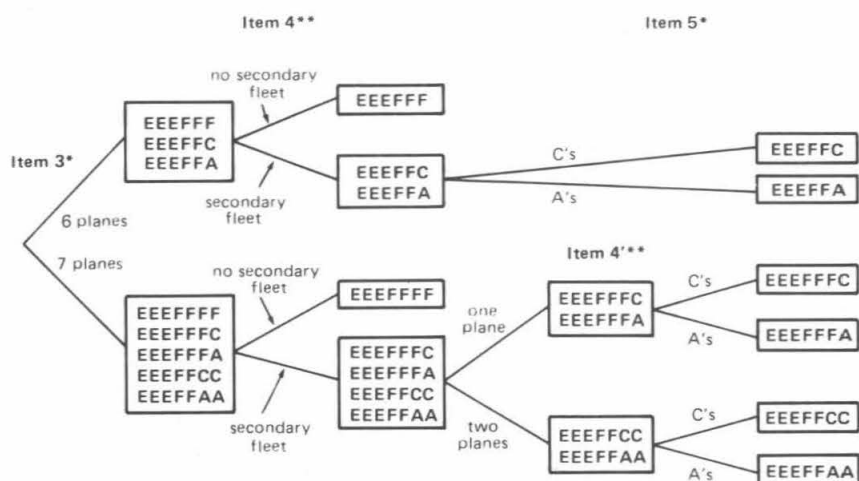


Figure 7.2. Schematic representation of flying club agenda. The item numbers marked with a single asterisk correspond to the numbers on the original agenda. The double asterisks indicate that the formal agenda listed only item 4, but the group correctly understood that it had two components—item 4 and item 4'. At the meeting the group did not vote formally on item 4, because no one advocated an unmixed fleet. The group simply moved to consider item 4' directly. From Levine and Plott (1977).

In order to test the effectiveness of our efforts, we designed a series of experiments. If the agenda failed to have an influence in a variety of experiments that involved conflicts similar to that of the flying club, we would be willing to say that our efforts had had no effect. If we found that we could use an agenda to influence group choice as we predicted, our confidence in the effectiveness of our efforts would be bolstered.

A questionnaire circulated by the club after the decision had been made provided data about the membership's preferences. The actual decision made by the group was checked against the prediction of the agenda model when applied to the reported preferences. The data were consistent with the model, but chance could still be an explanation. Next, we designed a series of experiments guided by the reported preferences. The objective was to see if the influence of the agenda could be demonstrated using those preferences in addition to preferences that would be even harder to manipulate.

Monetary incentives were included to induce preferences over an abstract set of options (letters of the alphabet). For example, an individual might receive \$8 if the letter A was chosen, \$5 if B was chosen, and so on. The amount a particular individual would receive

given the group choice of an option was known only to that individual, reflecting the fact that the happiness or degree of satisfaction of any member of the flying club was unknown to others. The induced preferences were similar to those reported by club members. The agenda was the same as the club agenda, except that all reference to airplanes and related terms was removed.

During the experiments, we demonstrated the influence of the agenda decisively by changing the agenda while holding preferences fixed. First an option was picked by the experimenter. The model was applied to find an agenda that would influence the group to choose that option. The experiment was conducted with that agenda. Then a second option was selected and an agenda was constructed to induce the group to choose it. An experiment was then conducted with that agenda. The agenda certainly and predictably influenced the decisions in these experiments.

Did the agenda influence the outcome of the flying club meeting? Two assumptions must be made in order to draw an inference: (1) The preferences of the flying club were similar to those induced in the laboratory. (2) The relationship between the act of voting and preferences was the same for the individuals in the flying club meeting as it was for those individuals in the laboratory. If both assumptions are accepted, the agenda must have had an influence on the outcome of the flying club.

The issue now focuses on assumptions (1) and (2). If (1) is doubted, new experiments can be conducted in which the induced preferences more closely approximate those hypothesized for the club members. In principle all preference patterns could be examined, so assumption (1) provides no problem for the application of experiments. Any criticism along the lines of (1) is not an objection to the use of experiments; quite the contrary, it is a call for more experiments.

Assumption (2) involves a theory of behavior. In essence it requires the hypothesis that the voting decisions of all people, including those of the flying club and those in our experiments, can be reasonably captured by the model used to design the agenda. The acceptance of this general theory is a key to the application. To the extent that it can be demonstrated to be unreliable, the conclusions drawn from the theory about the cause of the flying club decisions can be challenged. Again, however, we discover that the challenge does not involve an objection to the use of experimental method. Instead, the challenge is a call for additional theory and perhaps more experiments. Additional theory would be simply an improved replacement of the old, and the additional experiments would be tests of it.

The basic theory of the agenda appears to work well in the laboratory setting. The voting decisions of individuals depend on the packaging of options in the agenda. In the design of an agenda, this dependence can be relied upon to induce a majority to accept or eliminate a set of options. The extent to which one is prepared to assert something about the voting decisions of the flying club seems to be an unavoidable matter of subjective judgment concerning one's confidence in the two assumptions. A variety of preferences have been checked and, so far, no exceptions to the behavioral theory have been exhibited.

7.3 Demonstration: landing slot allocations

On occasion the implications of theory are so clear and the results of previous experiments so unambiguous that professionals have little to learn from experiments. Nevertheless, a theory that seems obviously relevant to professionals is frequently abstract and complicated to those who have the power and responsibility to make decisions. Sometimes in a policy-making environment even the word "theory" is pejorative. In such cases experiments provide a way of demonstrating the ideas without resort to theoretical constructions. The role of experiments as a demonstration of theory was the basis of a Polinomics report (Grether, Isaac, and Plott, 1979) on the allocation of airline landing slots.

After the Airline Deregulation Act of 1978, staff members of the Civil Aeronautics Board (CAB) became concerned about the method of allocating the right to land at four major airports (Washington National Airport, Kennedy, La Guardia, O'Hare). The allocation decisions were made by committees. Each airport had a separate committee consisting of those airlines that had been certificated by the CAB to operate at the airport.

In 1968 the Federal Aviation Administration (FAA) had limited the number of slots (takeoffs and landings per hour) that could be conducted at each of these airports. The committees were charged with the task of determining by "agreement" the allocation of slots among the certificated carriers. What might happen if the committees failed to reach unanimity was unclear. The FAA might have administratively allocated the slots, but the criterion it would have used was uncertain and the role of politics in the process made the consequences of default uncertain. Since the committees had successfully achieved agreement every six months from 1968 until the time of the study (1979), what might have happened if the committees had failed was only a matter of speculation.

The situation changed with the Airline Deregulation Act. The CAB

staff became concerned that the committees could be used as a barrier to new competition. I was contacted to study the committees because of my previous work on committee behavior. The question posed was related to the degree to which the committee process of allocating slots was compatible with the Airline Deregulation Act.

After we had undertaken some study and attended committee meetings, the nature of the committee behavior became clear and the structure of a reasonable process that could replace the committees became apparent. The appropriate model of the committee would have been immediately obvious to anyone with some game-theoretic and economic training. The committee operating under unanimity would attain some point in the "core" in the appropriate game without side payments. The location of the core would be very sensitive to the beliefs about what would happen if the committee defaulted. Simply put, an individual would rather veto an option than accept a committee decision less preferred than the option that would evolve as a result of the veto. The core would be an option preferred by everyone to the consequences of default. If such options existed, one of them would be chosen and, if not, the committee would default.

Since there were no side payments, an allocation in the core would not necessarily be efficient in an economic sense. That is, the airlines that would acquire the slots under the committee process would not necessarily be the carriers that valued the slots the most. In a cost-benefit sense the wrong carriers would get the slots. Furthermore, the allocation would be sensitive to airline beliefs about the consequences of default, and these beliefs would be sensitive to politics as opposed to economics.

Reasonable alternative processes involved markets with initial allocations determined by auction, or by lottery, or perhaps grandfathered with an aftermarket. Such alternative processes were very controversial and poorly understood by the airlines and public.

The role of the experiments was twofold. First, the experiments demonstrated the implications of the game-theoretic model used to evaluate the committee process. The emphasis is on the word "demonstrated" because the implications of the model were fully understood theoretically at the time, and previous experiments left little doubt that the substantive implications of the model would be predictive of committees operating under laboratory conditions. For those who had previously studied a wide range of committee experiments, very little was to be learned from additional experimentation.¹

¹ Consult Fiorina and Plott (1978). These experiments and subsequent publications provide substantial support for the core as a behavioral model under

The audience, which consisted of CAB staff, Department of Transportation (DOT) staff, FAA staff, and the airlines, had no previous experience with committee experiments. Nor did the audience understand or have a tendency to accept game theory. Thus given the political and controversial nature of the issue, some demonstration that the theory had content was necessary. The purpose of the experiments was to demonstrate the major consequences of the theory when applied to the slot allocation committee process while avoiding any detailed discussion of the content of the theory, the axioms, or mathematical structure and also avoiding any long and academic discussions about why the theory might be true. The strategy was one of demonstration.

The parameters for the committee experiments were chosen to reflect beliefs about the actual committee parameters. Part of the study involved a demand-curve estimation for a certain period at Washington National Airport. These parameters, scaled down appropriately, were the ones chosen for the experiment. The subjects in most experiments were adults, preferably with some connection to the aerospace industry (e.g., aeronautical engineers). These decisions were made in anticipation of a claim that the committees under the field parameters would behave differently and/or that people from the industry are different from other people. In addition, participants in some committees made several decisions together, reflecting the fact that sequences of decisions are characteristic of the slot committees.

Each individual was given a monetary incentive to acquire units of an abstract commodity, which from the experimenter's point of view represented the slots. In some experiments individuals participated in isolated committees. In other experiments individuals participated in more than one committee, and the value of slots received from one committee was dependent on the decisions made by the other committee. Such preference interdependencies or complementarities among the choice variables of different committees represented interdependencies among airports. A carrier might not want a slot to take off from O'Hare if it did not have a slot to land at Washington National Airport. The values of slots varied substantially among participants. These differences represented the different levels of economic potential that characterized different carriers.

Each committee had a fixed number of units to allocate. The rule was unanimity. In some experiments each individual was given a (different)

majority rule. Less experimental work existed on the behavior of the rule of unanimity, so technically speaking the experiments did have something to add.

quantity that represented the number of units he or she would get if the committee was unable to decide in the allotted amount of time (usually one hour). In other experiments the allocation was decided randomly in the event of a committee default.

The experiment made three points: (1) The outcome of the committee process is sensitive to the consequences of default; this point was made by experiments with identical preference parameters but different default rules. (2) The committee processes with different committee meetings for different airports could not deal efficiently with interdependencies among the airports. (3) The committee process would be insensitive to profitability of carriers and thus not promote an efficient allocation of resources. This point was made along with (1) by the inducement of high values for slots for some participants and very low values for others. The experiment demonstrated that, except for the bounds placed on decisions by the consequences of default, the value placed on slots by participants was unrelated to the allocation chosen. The allocation chosen was governed by the consequences of default and not participants' values.

The results of the committee experiments were not controversial. All three points were clearly evident in the data. Under unanimity a great pressure exists for equality of distribution, and unless a large allocation could be protected by a guarantee of a large quantity in the event of a default, participants had difficulty keeping it. For example, large carriers that should grow according to the economics of the situation never did so and usually contracted under the committee process. Inefficient carriers that should leave the airport never did so under the committee because they had no incentive to leave. The experiments provided a means by which the consequences of the theory of the core for the allocation of airport resources could be communicated without reference to the theory itself.

The report proposed the creation of a market for slots to replace the committees. Slots were to be auctioned by means of a first rejected bid, sealed-bid auction with an aftermarket. Markets and auctions had received some attention in the trade literature. Almost uniformly authors predicted that disastrous consequences would follow if markets were used to allocate the slots. This literature provided an excellent background for the experiments. The following questions were posed: (1) How do the committees perform relative to the proposed market process, and (2) do any of the disastrous predictions made in the trade literature actually occur if auctions are used?

Specifically the market experiments were conducted to demonstrate (1) that "rampant speculation" does not occur, (2) that the values

placed on slots by the large carriers do not dictate slot prices because price is determined by the marginal buyer, and (3) that the problems poorly solved by the committee process would be solved more efficiently by a particular type of market process. All three points were clearly demonstrated by the experiments.

The report was adopted and promoted by the CAB. It was the subject of many public hearings and a notice of proposed rule making. The recommendations to replace the committee with a market process were very controversial, but the experiments were never criticized. In fact, the economic analysis was accepted in the sense that the critics chose to question the CAB's authority to implement such a scheme, and the tools used by critics to back up such claims were congressional and international political pressure.

The exact role of the experiments in this process is difficult to ascertain. The report was supplemented by detailed transcripts of three of the committee meetings. Quotations from these meetings were used to buttress the results of the theory and experiments. No doubt these were read carefully, and from these texts alone the logic of the theory could be detected. The experiments probably prevented certain types of claims from surfacing in policy debates and also gave confidence to governmental staff members who needed to support their views with data. Something other than pure theory was necessary.

Staff at the FAA were opposed to market policies from the beginning. They were certainly not convinced by the experiments and funded experiments from another group that they hoped would disconfirm our findings. The follow-up experiments conducted by another group were so complicated that no conclusions could be drawn from them. They attempted to use members of the industry who applied their own valuations brought from the field. In the sense of modern experimental economics, the study had no controls.

The recommendations of the Polinomics report were not implemented in 1979. However, attempts to implement variations of the recommendations have been made almost every year since then. The analysis of the committees has been almost completely accepted by all, including the airlines and even the FAA. The committees themselves began to deadlock by 1982. By 1984, the airlines had recommended that the committees be replaced with a modification of the Polinomics recommendation that I proposed as an alternative (*Aviation Daily*, 1983). This alternative grandfathered airlines at current slot holdings, created a market for slots, and provided that new capacity be auctioned. The FAA, which had assumed a leadership role in opposing all forms of market allocation, aggressively opposed this proposal in favor of its own plan to allocate slots administratively. In the fall of 1985 the

DOT issued a notice of proposed rule making to implement the proposal. That rule became law on April 1, 1986. Since then legislation has been drafted in both the House and the Senate that would reverse the rule.

7.4 Shifts in the burden of proof

Experimental data can influence the burden of proof in an ongoing policy debate. In this context experiments seem to be as much tactical as a means of gathering facts about the nature of the actual situation. The objective of the experiment is to establish the need for proponents of the other side of the argument to prove or disprove something before a policy discussion can proceed in their favor. Specifically, in the cases discussed in this section, proponents of the other side had made claims about a complex situation based on a very general economic model. The experiments were designed to check the accuracy of that model. If the model advocated because of its generality failed to be reliable in the simple case of the experimental markets, the burden of proof would presumably rest on the advocate to explain why it did not work. If a model is so general that it can be applied to some very complex situations, one would naturally expect it to be reliable in the simple situations. If the model performs sufficiently badly in the experiment, the burden is on the model's advocate to explain why the experiments were "special" or "different" from the complex case in which the model is supposed to work. Failing that, the generality of the model is in question and the application to the complex case is in doubt. Thus the experiments do not address the field situation directly. Rather, they address the theory that one side or the other has used to analyze the field situation.

The strategy is not foolproof. The side that should accept the new burden of proof might choose to ignore the results. Intellectual honesty notwithstanding, something must force the burden if the strategy is to work. The shift in the burden-of-proof tactic has been explicitly used in two studies. Some of the demonstration arguments used by Grether, Isaac, and Plott in the Polinomics airport slot study could be counted as a third instance of shift of burden-of-proof strategies.

7.4.1 *Inland waterways barge traffic*

Railroad companies were lobbying a high-level administrator to require barges to post rates with the Interstate Commerce Commission (ICC). The railroads argued that the public information feature of posted rates would make the industry more competitive, permit the railroads to

compete more effectively against the barges, and aid the small barge owners who were allegedly secretly being undersold by the large barge companies. The administrator was skeptical of the arguments but had no basis for expressing his skepticism. He commissioned an experimental study (Hong and Plott, 1982) that became the first attempt to apply recently developed experimental methods to an actual policy problem.

The inland waterways barge industry is complex. Traffic exists on both coasts and in the Great Lakes region. Much of the industry exists on the Mississippi River and its tributaries. A great variety of products are hauled with boats, and firms are specialized accordingly. The first task of the study was to isolate a portion of the industry that had minimal complications. That portion would serve as a model for the creation of a laboratory industry.

A small portion of the Mississippi River was chosen. Only dry bulk cargo was incorporated into the basic model. Dry bulk was the major product for this stretch of the river. Parameters from governmental studies, judgments by industry people, and judgments by the researchers were used to characterize that portion of the industry during the year for which the best data were available. A laboratory experiment was conducted that represented a dramatically scaled down version of the industry.

The industry had about 15 grain shippers, the buyers of barge services. All were of approximately equal size. Between 25 and 35 barge companies existed. The size of a company could be measured by the number of boats it operated. Rough estimates of the volume of cargo shipped were available and served as the basis for demand and supply estimates.

The difficult part was determining an appropriate scale. Since a tow down the river took about a month, the number of boats translated into the number of tows a company could undertake. A unit in the experiment became one-half tow, and a period represented two weeks. A participant with a capacity to deliver five units in the experiment represented a company with more than five boats. The costs associated with units for sale in the experiment corresponded to engineering cost estimates for barges. An upward-sloping supply curve reflected a high marginal cost of upgrading marginal equipment and entry into the grain-hauling business by firms ordinarily hauling something else. The overall elasticity of supply was structured to reflect the best guesses about the industry. Elasticity of demand was similarly chosen. When the scale parameters were applied to the experimental parameters, the known industry aggregates were recovered.

A price-posting institution was used for two experimental sessions. The price-posting institution previously studied in laboratory work has many features similar to the rate-posting procedures used by the ICC. A second two experiments were conducted with a telephone market. All buyers and sellers were located in different rooms with telephones, and orders were negotiated and placed by phone. The telephone market was arguably analogous to the existing form of organization. Parameters were identical across all four markets.

In the price-posting markets, prices were higher, efficiencies were lower, and the small sellers made less profits than in the telephone markets. The results were exactly opposite to the predictions made by the railroads. Furthermore, the experiment provided estimates of the amount of business that would be shifted to the railroads if posted prices were required. The study was sufficient to make the administrator wary of the claims of the railroads. In private conversations they were challenged to explain the results. The administrator claimed that with evolving scientific evidence against their case he was not in a position to help them. The lobbying effort was diminished, and the policy advocated by the railroads was never pursued. The fact that a presumption existed against their case was sufficient to deter further lobbying efforts.

The administrator's use of this study was not widely supported within DOT. After the administrator left, the study was to have been published but was blocked by a staff economist who feared it would earn Proxmire's Golden Fleece Award and who in any case thought the idea of doing laboratory experiments in economics was foolish. At the time (1977) one could not use the authority of a large number of published papers to contest his belief. The study itself was then rejected by the *Journal of Political Economy*, which suggested that a paper with the details of the barge application removed and replaced by survey-oriented material might be acceptable. This suggestion came after the referees had mistakenly claimed that the results were due to an artifact of the experimental procedures.

7.4.2 The Ethyl case

The Federal Trade Commission (FTC) brought action against the major manufacturers of tetraethyl and tetramethyl lead, the lead-based gasoline additives that increase octane levels. The basis of the lawsuit was four practices widely used in the industry: (1) delivered pricing, (2) most favored nations clauses, (3) automatic matching of competitors' prices, and (4) advance notification of price increases. The experiments

reported in Grether and Plott (1984) were conducted for the FTC to be used as rebuttal testimony in this case.

The government's claim was that these four practices when taken as a group increased prices in an "anticompetitive" fashion. The logic is as follows. Delivered pricing removes the potential for under-the-table price discounts in terms of free services. Delivery is the only major service provided to customers by producers. Most favored nations assure customers that no other customer is buying at a lower price. This policy eliminates a source of small price concessions in response to individual customer pressure. It is similar to the posted price of a rate bureau. What you see is what you get – there are no negotiations. Meet or release clauses tie prices to that of a competitor. It is the precommitment to a trigger-price policy. A company will not win customers away from a competitor by price concessions because as soon as the lower price becomes known the competitor lowers its price automatically. Advance notification requires a 30-day notice in advance of price increases. By giving a 40-day advance notice, competitors were aware of a 10-day window to bring prices up to the new level. If they failed to act in 10 days, the company that made the notice would necessarily retract it because of the practice of matching prices. Thus by giving a 40-day notice a company gives competitors a choice between all competitors having the higher price and there being no price increases by anyone.

The four defendants' reply to the charge that the practices had an anticompetitive effect was that they were an oligopoly. According to the defense, the practices had no effect on industry performance because there was no room for an effect. The industry enjoyed high (but not illegal) prices fostered by industrial concentration. According to the defendants' claim, any high prices and/or profits were accounted for entirely by industrial structure and therefore were unrelated to the four practices. The profits were not eroded by entry because the government's decision to phase out leaded gasoline served as an effective barrier to entry.

The experiments were designed by Grether and Plott to serve as a basis for rebuttal testimony for the government. Was the industry's claim true? Is it a fact that the practices necessarily have no influence when the industrial organization is that of the industry? If the answer is no, the defendants could not claim that the high profits and prices are necessarily unrelated to the government's case. A major tenet of their argument would be damaged.

Laboratory experiments were designed to match the numbers of suppliers and demanders, concentration ratios, demand elasticities,

excess capacity, and so on, that are known properties of the industry. Special care was taken to anticipate questions that one could imagine during cross examination. Would the attorney attempt to make the experimental argument look foolish? The subjects were not undergraduates. For the most part they were employed adults preferably with an engineering background and/or some connection to the oil industry. The subjects participated in more than one experiment. Several different variations of the practices were studied. Consistency in design with that of previous experiments was sought so the weight of the authority of different types of experiments conducted by others could be used. Wherever possible, the consistency of the results with "the tradition of experimental research" was established. Testimony of respondents' experts was studied carefully so different forms of the rebutted theory would be recognizable within the experimental design. Many replications were done. Some experiments were blind in the sense that the experimenter conducting the experiment did not know the parameters. A double blind experiment was considered, but the experiment was so complicated that it could not be conducted by novices.

The results of the experiments were decisive in showing that the practices could have a substantial impact. The results were circulated to the respondents, but the government decided (correctly) that the case could be won without the rebuttal testimony provided by the experiments. Since these experiments were novel and since experiments had never before been introduced in any court, a decision was made not to enter the experimental results into testimony. Presumably there would have been no problem getting the experiments admitted as evidence (Kirkwood, 1981). The government won the case on the first round, but the defendants won on appeal.

After the trial a seminar on the experimental results was conducted at the FTC. Discussions with the defense lawyers following the seminar revealed some of their thinking. They had studied the several variations of the practices reported in the paper. One of the variations in which the practices were not strictly enforced resulted in prices slightly above the competitive equilibrium. Counsel for the defense asked if that treatment was the best approximation of the actual practices. The questions made sense because the practices as found in the industry were not perfect. Evidently, the first line of defense would have been to attempt to use the experimental data as evidence in support of the counsel's own position. In retrospect our experimental case could have been a better tool for the prosecution if we had built the detailed exceptions to perfect enforcement into the design. Given

the nature of the imperfect practices found in the field, the experimental results would probably not have changed at all had this change been made.

An interesting feature of all three attempts to use the shift of burden of proof strategy is that the experiments were designed to mirror the industry as closely as possible. Relative sizes of buyers and sellers, demand elasticities, numbers of participants, and so on were all similar to those of the target industries. This was done to prevent the application of theories that attempt to show that the behavior of the laboratory industry will differ from that of the industry. Each imaginable difference between the experimental setting and the field is the starting point for a potential theory. An infinite number of such theories necessarily exists.

The logic is as follows. Individual A (railroads in barge study, the defense in the Ethyl case) has used a general theory T to infer something about the industry and its performance. Individual B (the experimenter) has noted that, under experimental circumstances E, theory T is not reliable. Thus T is not reliable in general because it is not reliable in E. Individual B then asks A to explain why T can be applied to the industry. That is, B places a burden of proof on A to show why T is applicable to the industry but not under condition E. Now B does not want A to have readily available some specified property of E that might be used to argue that E is an (uninteresting) exception to the general reliability of T. Each difference with the industry serves as a basis of a potential theory. For example, the laboratory results might be attributed to the special concentration ratios used in the experiment that differ from the industry's. The laboratory results might be attributed to the use of people experienced in the industry. The laboratory demand elasticity might differ from the industry, and so on. Such theories can be checked through additional experiments, but time and money are involved. The best strategy is to eliminate as many potential theories as possible so the burden of proof is not easily shifted back to its original position.

7.5 Direct extrapolation: air freight posting

Policy choices require making decisions, and the weight of the evidence is a subjective issue that rests with the decision maker. Studies designed to answer one set of questions might provide a decision maker with sufficient insight to act on a completely different set of issues. Such was the case with the CAB air freight rate decision.

Before 1980 air freight forwarders were required to post their rates with the CAB. Having studied the influence of posted prices in the

early Plott and Smith (1978) study and the barge study (Hong and Plott, 1982), the CAB made the reasonable presumption that posted prices reduce market efficiency. On the basis of existing laboratory results and in the spirit of deregulation, the CAB issued a notice of proposed rule making to eliminate air freight rate posting. Seeing no claims that the presumption was incorrect, the CAB acted and eliminated the policy of rate posting.

7.6 Potential design: prepolicy research

Two experimental studies (Plott and Wilde, 1982; Lynch et al., 1984) were developed *as tools* with which to study policy options. Both were initially financed by the FTC, which has an interest in consumer protection. The staff of the commission is exposed to many competing policy prescriptions aimed at correcting alleged market failures.

The problem faced by the staff was that neither the "market failure" nor the influence of a "proposed remedy" can be clearly observed with field data. The experimental strategy was to create markets that would reliably "fail." Such markets could then be used to study the properties of proposed policy remedies as implemented in those markets. The experiments conducted were not focused on any particular industry or potential decision, nor were they designed to "test" any particular theory directly. Rather, they had characteristics of a variety of markets and alleged market problems that were the concern of the commission. The degree to which some theory or model might help explain their behavior was a secondary concern.

The experiments were complicated. The use of random devices and the associated training were nearly a separate experiment. Because of the nature of information acquisition and use, new types of market organization were imposed. Several different types of institution were studied that sometimes varied according to subtle aspects of when people were informed about their own preferences, the properties of commodities, and what they might do as a consequence. A full description of these experiments is far beyond the scope of this chapter.

The Plott and Wilde study focused on markets in which the consumer has only a limited capacity to evaluate the commodity (e.g., the services of a physician or perhaps the services of an automobile mechanic) and in which the seller might have a financial incentive to mislead the buyer. Experiments were conducted in which sellers offered two types of commodities (e.g., medical procedures). The relative value of these commodities to buyers depended on the state taken by some unobserved random variable (e.g., the infirmity). Only

clues or statistics dependent on the state (e.g., symptoms) were observable. In some experiments the buyer could observe and interpret the clues. Sellers in these markets had a complicated supply response because of the multimarket nature of the setting, but they had no special function of information supply. When the sellers could interpret the clues and the buyers could not, the sellers had an additional important function. In addition to a supply response, the sellers gave advice to buyers and interpreted the buyer symptoms as part of the sale. The economic questions focused on models of the efficiency with which such markets operate. Would the buyer receive good information? Would the buyers act on it?

The Lynch et al. study also focuses on asymmetric information in markets. Markets were created in which low-quality products were delivered even though high-quality deliveries were Pareto superior. In this sense the study was very successful, because we were able to create such markets. The study then turned to the role of reputation and warranties in improving the efficiency of the markets.

An interesting feature of the Lynch et al. study is an experimental design involving sequential decisions. Recall that "theory tests" were not the primary purpose of the experiments. Instead, the strategy was first to try an extreme case in which almost all models predict that "lemons" would be produced. If the lemons phenomenon was demonstrated by the first phase of experimentation, the strategy was then to see if the most likely corrective policy as suggested by most theories would be effective. This policy involved the introduction of required, costlessly enforceable warranties. A failure at either end of this spectrum would have provided the background for a large series of negative results (about the applicability of theory) and subsequent experiments of a completely different sort. The positive results at the extremes that were actually observed were used as a foundation for exploring the more delicate phenomena between the extremes in which warranties were not required, enforceable, and so on.

Neither proposed legislation nor proposed rule making resulted from either study. The researchers learned much about the limitations of broad statements concerning behavior that have accompanied past policy decisions. The experiments also provided many insights into the nature of models that are being applied to consumer protection problems. The hope is that these background experiments will be the basis for additional experimentation and policy analysis.

7.7 Design

Sometimes economic problems require completely new types of organization and decision processes. In such cases history supplies no data,

and a unique opportunity exists for experimental work. Experiments can provide some, perhaps limited experiences on which to base judgments about the nature of appropriate organizations and policy. Such situations are properly called problems of organizational design. Three instances are outlined here.

7.7.1 *Slot exchanges*

Following the slot allocation process study by Grether et al. (1979), the air controllers' strike occurred and the committees at the various airports began to deadlock. The number of constrained airports expanded from 4 to 22. A decision was made to create a "slot exchange." Air carriers were given temporary grandfather rights to their historic level of slots. The plan was to allow carriers to meet and exchange slots within and across airports on a one-for-one basis.

How should the process be organized? The problem was nontrivial because the size of the exchange was staggering. The number of commodities measured in the thousands, and the number of agents measured in the hundreds. The politics of the situation dictated that no buying or selling be allowed, so no numeraire existed. The logistics problem was enormous.

My role in this process was as a consultant for a carrier that wished to trade away from one airport to get to another. As a participant in all organizational meetings, I was involved in the design of the process.

The ultimate process was constructed on the basis of experiences with experiments with one-sided oral auctions. The only difference was that bids were to be tendered in writing rather than orally. Each carrier listed slots that it wished to acquire together with slots that it would take in exchange. The form of these proposed trades was that any slot in column A would be exchanged for any slot in column B. These lists of bids were collected and circulated to all carriers. With the list of proposed trades, carriers searched for chains of trades that involved their own proposals. At this stage of searching for trades, carriers were free to add new proposed exchanges that had the effect of a proposal being accepted or a chain of trades being completed.

The process was not well understood. However, pilot experiments had been conducted at the California Institute of Technology. The carrier that had employed me had practiced and had a strategy for dealing effectively with the process. Since our "team" had well-defined ideas about how the logistics of the process might work, we had little difficulty in convincing the group of all carriers to adopt the process. The process and improved variations were used many times, including a brief period when monetary transactions could take place

and a period in which "many-for-one" trades were permitted. The entire affair was similar to a large experiment, and given the constraints it worked rather smoothly.

7.7.2 Westchester County Airport

The county of Westchester in New York decided to auction access to its airport terminal. The terminal is small and safety codes limited passenger usage to a maximum of 40 enplane and 40 deplane passengers in any 15-minute period. In addition, a maximum of four aircraft could use the parking pads at any time, and at most, two of these could be large aircraft.

When the county had taken action to limit the use of the terminal to the stated capacity limitations, it became involved in a lawsuit. The judge ordered the county to devise a mechanism for allocating the available capacity that was consistent with the Airline Deregulation Act. The county chose to develop an auction system that was to be used in the event that a settlement could not be attained.

The auction was designed by Glen George, a graduate student at the California Institute of Technology, and myself. It was important to avoid many potential criticisms of auction processes that litigants might raise, and it was necessary to conform to the realities of politics. The carriers might more readily accept a process that tended to allocate rents to carriers, so a one-price auction was used. In a discriminative auction, sellers pay what they bid. In a one-price auction, the high bidders pay the value of the excluded bid. If the demand curve is "steep," the former generates greater revenue to the seller than does the latter (Miller and Plott, 1985). Because the continuum of time was unwieldy, the day was divided into 15-minute segments. Capacity was divided into five passenger enplane blocks and five passenger deplane blocks. Thus two separate markets existed every 15-minute period of the day in which eight passenger blocks were sold in each. Carriers wished to tie purchases together, so provisions for special constraints that tied enplane purchases to deplane purchases were designed. Carriers were also allowed to submit multiple bids tied together with a constraint that canceled all other tied bids if one of the set was accepted.²

The number of markets together with the possibility of constraints resulted in a very large and potentially complex auction. Experiments

² Many of the ideas were motivated by Rassanti, Smith, and Bulfin (1982). The combinatorial auction was not feasible because of practical and technical problems.

were conducted using demand conditions similar to those believed to exist at Westchester County Airport. The purpose of the experiments was to answer some very practical questions: (1) Were the instructions clear about how to tender bids and use the constraints? What types of confusion were we likely to encounter? (2) Did unusual strategies exist that might undermine any efficiency properties of the auction process? (3) Were we likely to encounter a computational problem in determining the winning bids? We could imagine problems that would exceed our computer capacity. The solution to the auction involves a large integer programming problem, the dimensions of which are very sensitive to the number of constraints. The use of bids and constraints is not governed by the logic of the problem, so we had no a priori way of determining the size of the computational problem without actually trying the auction.

The experiments were invaluable. Many problems were uncovered at every stage. The instructions were not clear. We did not have a firm grasp of the game theory, and the experiments provided insights about the strategic possibilities. Computational problems did exist. The whole process was redesigned several times after bugs surfaced during the experiments. Experiments are still being conducted to improve the process.

The process was not used at Westchester County Airport. The respondents settled by adopting the process I proposed for Washington National Airport (*Aviation Daily*, 1983) discussed in Section 7.3. The original laboratory experiments and the field experiments with the slot exchanges have provided convincing evidence that markets in slots will "work." All of this evidence made carriers happy with a market for slots, although the FAA remained adamantly opposed. Now the New York Port is considering an auction process for the three major airports in New York City. The research and refinements on the Westchester County problem are relevant to the Port's problem.

7.7.3 *Space station*

The National Aeronautics and Space Administration (NASA) is planning to place a manned space station in orbit. The station will be used as a research laboratory, as a manufacturing facility, and for a variety of other purposes. The users will be the U.S. government, foreign governments, and private corporations. The Reagan administration wants the capacity to be allocated by some sort of pricing system. A team of economists at the Jet Propulsion Laboratory has been given the task of designing a pricing mechanism.

The task is complicated by the existence of nonconvexities, externalities, large common costs, much uncertainty, and other factors. In addition, NASA cannot operate for a profit or even take money for that matter, so profit centers and related decentralized schemes do not seem to be feasible. Matters are further complicated by the fact that the space station design is in a continuous state of evolution, and the design of the system itself should be affected by the pricing scheme.

The proposed role of experiments in this project is much different from that in previous research. Some testing of institutional influences is underway (Banks, Plott, and Porter, 1986), but the central role is to be slightly different. The experiments are to be used as a heuristic – a tool for organizing thoughts and questions as opposed to a tool for answering questions. Most experimenters have noticed that the process of designing experiments makes the researchers aware of complications and interdependencies that would have otherwise escaped notice. The space station project is intended to capitalize on this feature of the method.

The space station is just finding its way to the drawing boards. The variables are not even known, much less the costs. The experimental plan is to conduct simulations of pricing policies under experimental conditions that reflect much of the physical, institutional, and motivational aspects of the space station. The ultimate subjects will be the NASA personnel who are building the station. The purpose will be to instruct the personnel on the nature of competing policy options by providing them with some experiences with their operation. It is hoped that such exercises will generate insights about the features of the station, its cost, engineering structure, service capacities, and so on and thus make the simulation of policy options useful.

7.8 Closing remarks

The theme of this chapter is that “parallelism” involves many different dimensions that reflect the nature of policy analysis. Parallelism takes different forms, but there seems to be no formula for choosing an appropriate one. Instead, the use of laboratory methods in policy contexts seems to be an art involving a skillful and very subjective choice of experimental conditions. The laboratory results are sources of experience under conditions that it is hoped will be useful to those responsible for making decisions. The role of experiments in policy contexts is an activity more akin to practice than to some sort of scientific pursuit of truth. Experimentation is a source of experience similar to the experience one acquires as one practices the piano before

a concert or that a team acquires as it practices before a game. The connections between such experiences and final performance can have many dimensions. The kinds of useful practice undertaken in preparation for a football game can range from ballet to scrimmage. The scrimmage itself can involve plays that the coach thinks the opponent will use as well as plays the coach is sure will not be used but are nevertheless educational. Similarly, experiments that provide the best insights about the nature of upcoming options might include faithful reproductions of the anticipated situation, but certainly there is no reason to believe that good experiments necessarily take that form.³ In fact, there is no *a priori* reason to believe that faithful reproductions would be of any use at all.

Having compared this type of research with practice, it may come as no surprise that I am particularly uncomfortable with the concept of external validity. First, the word "external" involves a needless commitment to the proposition that there are no general principles of behavior that govern simultaneously both laboratory and "other" situations. If both laboratory behavior and field behavior are governed by the same principles as is believed to be the case in economics, it makes little sense to think in terms of "external" and "internal" behavior; all of the behavior is "internal." Second, the word "validity" sets a standard that is impossible to meet in policy contexts. An economic policy decision will constitute a unique event in history. The exact circumstances will not be repeated. Many unobserved parameters will be in operation. There is in principle no way to "validate" theories about what might happen. Simple judgment cannot be avoided. The experiments simply shape the thought processes, the data, and the arguments that form that judgment.

I do *not* intend to suggest that policy applications of experiments involve only rhetoric. Although opinions of policymakers are clearly important, the research objective is not simply to alter opinions. The objective is to make a correct guess about what will happen if a policy is put into operation. The purpose of the experiment is to make the guess as informed in the light of experience as possible.

³ My own thoughts about how one might learn something from experiments were influenced by environmental engineers at the California Institute of Technology. The engineers were attempting to determine the flow of effluents that might result from a change in release in the ocean near Los Angeles. They were studying currents in a large pool constructed in the basement of a building on the campus. Of course, their pool looked nothing at all like the Pacific; yet it taught them something about their models, and it was the models that helped them learn something about the Pacific.

My approach to applied work has been to forget the concept of external validity and not to take the concept of parallelism too literally. Instead, the approach has been pragmatic in the sense that the use of experiments in each project has been justified by whatever arguments seemed appropriate given the context. What kind of experiment would make the guess work inherent in policy decision making more informed? The purpose of this chapter was to survey some of what was done to see to what extent some order or method actually existed after all. Laboratory studies and policy problems are connected by a many-dimensional correspondence. This chapter has outlined a few of those dimensions.

References

- Aviation Daily* (Washington, D.C.). July 25, 1983, back of pp. 124 and 127.
- Banks, Jeffrey S., Plott, Charles R., and Porter, David P., "An Experimental Analysis of Public Goods Provision Mechanisms with and without Unanimity," Social Science Working Paper No. 595. California Institute of Technology, January 1986.
- Fiorina, M., and Plott, Charles R., "Committee Decisions Under Majority Rule: An Experimental Study." *American Political Science Review* 72 (June 1978), 575-98.
- Grether, David M., Isaac, R. Mark, and Plott, Charles R., "Alternative Methods of Allocating Airport Slots: Performance and Evaluation," prepared for the Civil Aeronautics Board. Pasadena, Calif.: Polinomics Research Laboratories, Inc., 1979.
- "The Allocation of Landing Rights by Unanimity among Competitors." *American Economic Review* 71 (May 1981), 166-71.
- Grether, David M., and Plott, Charles R., "The Effects of Market Practices in Oligopolistic Markets: An Experimental Examination of the Ethyl Case." *Economic Inquiry* 22 (October 1984), 479-507.
- Hong, James, and Plott, Charles R., "Rate Filing Policies for Inland Water Transportation: An Experimental Approach." *Bell Journal of Economics* 13 (Spring 1982), 1-19.
- Kirkwood, John B., "Antitrust Implications of the Recent Experimental Literature on Collusion." In *Strategy, Predation, and Antitrust Analysis*, edited by Steven C. Salop, pp. 608-21. Washington, D.C.: Federal Trade Commission, September 1981.
- Levine, Michael E., and Plott, Charles R., "Agenda Influence and Its Implications." *Virginia Law Review* 63 (May 1977), 561-604.
- Lynch, Michael, Miller, Ross M., Plott, Charles R., and Porter, Russell, "Product Quality, Informational Efficiency and Regulations in Experimental Markets," Social Science Working Paper No. 518. California Institute of Technology, March 1984.
- Miller, Gary J., and Plott, Charles R., "Revenue Generating Properties of Sealed-Bid Auctions: An Experimental Analysis of One-Price and Discriminative Processes." In *Research in Experimental Economics*, edited by Vernon L. Smith, vol. 3, pp. 31-72. Greenwich, Conn.: JAI Press, 1985.
- Plott, Charles R., and Smith, Vernon L., "An Experimental Examination of Two Exchange Institutions." *Review of Economic Studies* 45 (February 1978), 133-53.
- Plott, Charles R., and Wilde, Louis L., "Professional Diagnosis Versus Self Diagnosis: An Experimental Examination of Some Special Features of Markets with Uncer-

- tainity." In *Research in Experimental Economics*, edited by Vernon L. Smith, Vol. 2. Greenwich, Conn.: JAI, 1982.
- Rassanti, S. J., Smith, V. L., and Bulfin, R. L., "A Combinatorial Auction Mechanism for Airport Time Slot Allocation." *Bell Journal of Economics* 13 (Autumn 1982), 402-17.
- Samuelson, Paul A., and Nordhaus, William D., *Economics*, 12th ed. New York: McGraw-Hill, 1983.
- Smith, Vernon L., "Relevance of Laboratory Experiments to Testing Resource Allocation Theory." In *Evaluation of Econometric Models*, edited by J. Kmenta and J. Ramsey. New York: Academic Press, 1980.